

Problema 2:

- a) Plottando il grafico dei punti, è evidente come i gruppi del dataset siano 3. (guarda fig. 2)
 Anche il dendrogramma del clustering con Ward ci conferma questa ipotesi: si evince chiaramente (vedere fig. 1 per il dendrogramma) come si formino 3 clusters.

$$\Rightarrow \boxed{K=3}$$

b) Assunzioni:

$$\left. \begin{array}{l} \underline{X}_{11} \dots \underline{X}_{1n_1} \text{ iid} \sim N_p(\underline{\mu}_1, \Sigma) \\ \vdots \\ \underline{X}_{31} \dots \underline{X}_{3n_3} \text{ iid} \sim N_p(\underline{\mu}_3, \Sigma) \end{array} \right\} \perp$$

- 1) Gaussianità dei sotto gruppi $\{1, 2, 3\}$ ✓
 \Rightarrow confermata dal plot MC test, in tutti e tre i gruppi p-value molto alti.

- 2) stessa matrice delle covarianze Σ . ✓

In questo caso non esiste un test per Σ multi-dimensionale. Qualitativamente (plottando S_1, S_2, S_3) vediamo che le sample covariances sono molto simili. (fig. 3)

\Rightarrow Possiamo procedere con il test. Utilizziamo la statistica di Wilks per MANOVA.

$$\text{p-value del Test: } \approx 0 < 2.2 \times 10^{-16}$$

$$\Rightarrow \text{Rifiutiamo } \boxed{H_0: \underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3}$$

c) CIs di Bonferroni:

Ci occorrono 6 test: $\left\{ \begin{array}{l} 3 \text{ differenze} \\ \times \\ 2 \text{ dimensioni: } \{ \text{length, width} \} \end{array} \right.$

$$\Rightarrow K=6$$

$$\text{quantile: } t_{\frac{\alpha}{2K}}(n-g)$$

$$g=3$$

• Differenza gruppo {1, 2}:

Length: $[-0.222, 0.147]$

Width: $[-3.182, -2.473]$

• Differenza gruppo {1, 3}:

Length: $[2.07, 2.41]$

Width: $[-1.19, -0.55]$

• Differenza gruppi {2, 3}:

Length: $[2.08, 2.47]$

Width: $[1.58, 2.32]$

Conclusioni:

- I gruppi {1, 2} hanno statisticamente la stessa lunghezza^(media), mentre il gruppo 2 ha ampiezza ^{media} maggiore del gruppo 1.
- I gruppi {1, 2} hanno lunghezza ^{media} maggiore del gruppo 3. Per quanto riguarda l'ampiezza, il gruppo 3 è più ampio rispetto al gruppo 1, ma meno ampio ^{in media} rispetto al ~~gruppo~~ gruppo 2.